

Empirical consequences of definitions: the case of unemployment in German register data

Kruppe, Thomas

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Kruppe, T. (2009). Empirical consequences of definitions: the case of unemployment in German register data. *Historical Social Research*, 34(3), 138-148. <https://doi.org/10.12759/hsr.34.2009.3.138-148>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Empirical Consequences of Definitions. The Case of Unemployment in German Register Data

Thomas Kruppe *

Abstract: »Definitionen und ihre empirische Folgen – „Arbeitslosigkeit“ in deutschen Register Daten«. When using national register data for quantitative analysis, researchers depend on the quality of the data provided by the institutions collecting the data. Data quality is strongly influenced by institutional settings during data collection, e.g. underlying measurement concepts. Using the example of the duration of unemployment in Germany, this paper discusses the effects of the definition and their implementation on data quality. Therefore, two different definitions of (registered) unemployment are discussed. Implementing them into register data derived from German Public Employment Service, the effects of classifications on statistical results are shown.

Keywords: Longitudinal Analysis, Process-Generated Data, Social Bookkeeping Data, Public Administrative Data, Institutional Filters, Measurement, Unemployment, Labour Market Data.

1. The Problem of Changing Measurement Concepts

Emergency of Unemployment – and how to combat - is a big issue in social research. Studies focusing on this topic vary data sources as well as the operationalisation of unemployment. The latter is because unemployment inherently is a concept or, moreover, a social category, which is highly related to the more general concept of a labour market with status categories like employment, inactivity, retirement or being out of the labour market.¹

In order to illustrate the effects of changing definitions, this paper will start with two contrasting definitions of unemployment: Section 2.1 directs attention to the national definition and institutional settings of registered unemployment in Germany (definition 1), while Section 2.2 discusses an alternative concept of unemployment (definition 2). The rest of this paper will test the effects of an implementation based on these definitions on statistical analyses, using register data from the German Federal Employment Service (FES)². An overview on the structure of register data is given in Section 3. The empirical implementa-

* Address all communications to: Thomas Kruppe, Institute for Employment Research (IAB), Regensburger Str. 104, 90487 Nuremberg, Germany; e-mail: Thomas.Kruppe@iab.de.

¹ On the genesis of unemployment as social problem and the formation of social institutions see Zimmermann (2006). On the concept of labour markets as a system of individual status transitions see Kruppe (2003).

² The FES in Germany is named Bundesagentur für Arbeit (BA).

tions of the two definitions into register data is demonstrated in Section 4: The duration of unemployment most closely approximate to the calculation by the FES (implementation 1) is contrasted by the alternative concept based on the discussion in Section 2.2 (implementation 2). Empirical evidence is shown that there is a clear impact on results depending on the underlying definition of unemployment.³ This is highlighted in the conclusions (Section 5).

2. Differences in Definitions of Measurement Concepts

Nowadays the term “unemployment” is very common. But asking for a definition, a broad spectrum of answers are given. The number of unemployed can be derived from the register of the FES⁴. But the definition varies over time and – more important from a sociological perspective – may not meet the theoretical approach of a researcher. To give an example, section 2.1 will roughly introduce the legal definition of unemployment in Germany (definition 1). In contrast, Section 2.2 interprets unemployment more general as not being in employment (definition 2), as it could be derived from theory of social exclusion.

2.1. National Definition and Institutional Settings of Registered Unemployment in Germany

German standards of unemployment are defined by legislation. To be precise they are codified in the Third Volume of the Social Code (Sozialgesetzbuch III (SGB III)) and the Second Volume of the Social Code (Sozialgesetzbuch II (SGB II)). Both could be seen as institutional settings of the German social security system and specific characteristics of the welfare state (see Esping-Andersen 1990). But while the SGB III is codified from a labour market perspective, the SGB II originates from a welfare perspective for those who are needy. Therefore the SGB II broadens the definition of unemployment to all individuals who are needy as well as capable of working.⁵ These legal definitions of unemployment vary on a number of dimensions, for example:

- 1) *Individuals vs. Households*: In the case of SGB III unemployment is part of the more general concept of a labour market with individual status catego-

³ For an international comparison and an enhanced description on 64 ways of calculating unemployment see Kruppe et al. (2008).

⁴ Alternatively, the number of unemployed could also be estimated from survey data. While in the first case (register data), the definition is set by legislation (eligibility), in the second case (survey data), among others the definition is depending on the wording of the survey.

⁵ The second title of the SGB III, introduced in 1998, is “Arbeitsförderung” (literally translation: “Labour Promotion”), including regulations about eligibility and individual calculation of unemployment benefits. The second title of the SGB II, introduced in 2005, is “Grundsicherung für Arbeitssuchende” (literally translation: “Base security for jobseekers”), introducing a unified benefit system for needy employable people.

ries like employment, inactivity, retirement or being out of the labour market. In the case of the SGB II, unemployment is a social category related to income and means - not only of the person affected, but the household the individual is living in.

- 2) *Eligibility to Unemployment Benefits vs. Registered as Jobseeking*: Therefore, the definition in legislation (SGB III) is generated focusing on the eligibility to register at the FES and coverage of unemployment benefits as part of the social security system. It starts with the distinction between persons in or not in employment, jobseekers, unemployed persons, disabled and re-entrants:⁶ From social security perspective the focus is put on registering as unemployed, in contrast to registering as a jobseeker with only specific entitlements (see below). Nevertheless there are still overlaps between definitions.
- 3) *Employment Status vs. Work Hours*: The status of employment can be interpreted as self-employed, marginally employed, family work or – most important – dependent employed with compulsory social insurance contributions. From social security perspective persons are counted as *not in employment*, if they are a) not employed at all or b) are only employed for less than 15 hours per week. *Jobseekers* are defined as persons looking for employment with a duration of more than 7 calendar days. They must register personally at the local agency of the Federal Employment Service for placement, must be capable and allowed to pursue the achieved occupation and to reside in Germany. To register as a jobseeker, it is not necessary to be registered as unemployed. To *register as unemployed* one is additionally obliged to search for employment of at least 15 hours with compulsory social insurance contributions of more than 7 calendar days (see definition of jobseekers below). In this context “to search” means that the person has to strive any possibility to end the period of having no job (of at least 15 hours). This includes being at the disposal for a placement through the Federal Employment Service. Therefore, availability to work is defined as to be capable of work and being willing to take up any employment offered immediately. To check for, one has to register personally at the FES, to show up regularly and to keep any appointment (committed number of application a week, taking part in an active labour market policy measure, etc.).

As a consequence, the legal definition of unemployed persons in the context of SGB III can be understood as jobseekers, who are temporarily not in employment but searching for employment with compulsory social insurance contributions of at least 15 hours and more than 7 calendar days, have registered at the Federal Employment Service personally and are capable of work and are willing to take up work immediately.

⁶ These definitions are explicit valid only in the legal context of the SGB III.

This definition is only partially connected to the facts of eligibility of unemployment benefits: On one hand, the criteria of receiving unemployment benefits is not sufficient to be counted as unemployed. This is, for example, the case for persons who cannot take up work because of illness up to six weeks. On the other hand, one can be registered as unemployed without receiving unemployment benefits (disabled person, women re-entering the labour market after a period of child care, etc.). Implicit, persons are not counted as unemployed who are ...

- employed more than 15 hours a week,
- not capable to work,
- not allowed to work (like persons younger than 16 years),
- not available without reasonable causing,
- not registered personally at the Federal Employment Service, do not show up again for longer than three months, or did not keep an appointment for several times without cogent reason,
- taking part in active labour market policy measures⁷,
- unable to work because of illness (after six weeks),
- doing military or compulsory service or are arrested,
- pupils, students or school leavers, looking only for professional training / formation,
- foreigners without compulsory work permit and persons seeking / granting asylum, including the members of their family,
- on leave (i.e. parental leave, sabbatical)

Introducing the SGB II - and in contrast to former legislation - anybody who is in need of and wants to receive social benefits nowadays is categorised as being unemployed. This classification requires only the ability to work for at least three hours a day. The definition of unemployment was therefore broadened to anybody needy and capable of work, both interpreted in a household context for any person who could be obliged to pay alimony or palimony. This refers to concepts of labour reserve, hidden unemployment, hidden labour force and discouraged workers, while these are now categories depending on a reasonable maintaining in the household context.

This change in legislation has had a clear statistical impact on the official unemployment numbers which are significant higher from 2005 ongoing.

⁷ One exception were short training measures ("Trainingsmaßnahmen") with a duration between 1 day and 12 weeks, which had very heterogeneous contents. Persons taking part before 2004 were still counted as unemployed. This is no longer the case since then.

2.2. Alternative Concepts of Unemployment

The number of unemployed can be derived from the register of the FES, using the legal definition of unemployment in Germany, described in section 2.1 (definition 1). In contrast, this section exemplifies the variety of possible definitions interpreting unemployment more general as “not being in employment” (definition 2), utilising unemployment as being not attached to the labour market with its consequences on social exclusion.

The individual labour market status, and therefore the status of unemployment, is time-dependent. Additionally, different underlying definitions of labour market statuses can lead to different interpretations of individual employment histories,⁸ also in the individual number of phases of unemployment and their duration.⁹ While any interruption of unemployment is counted as an outflow from (begin) and an inflow into registered unemployment (end of the interruption), the duration is a retrospectively calculated sum of days the status matched the criteria discussed above. This section deals with the question, which periods could be seen as unemployment and therefore included into the calculation of its duration – independent from the legal framework of the unemployment register as part of the social security system.

Remembering the fact that individual employment histories are often interpreted as a major key for individual participation in modern societies (i.e. social in- or exclusion) as well as for future (employment) probabilities, the question of measurement is nontrivial. As an example, the status of unemployment could be interpreted as any period of not being in employment. Doing so, there are several labour market statuses that can be added under the definition of unemployment, for example

- participating in active labour market policy measures,
- illness or maternity protection,
- periods without legal basis to take up employment,
- interruptions without documentation.

Also other modifications in the concept could be discussed, depending on the theoretical framework of research is undertaken. Based on the general information available in German register data (see Section 4), a possible change in defining unemployment could be as follows:

Starting at a point in time of interest, anybody is defined as unemployed if registered unemployment is shown by the administrative record. These individuals are fulfilling the rules of the legal framework (concept 1).

⁸ For an example see footnote 5.

⁹ On the general underestimation of long-term unemployment when using register data see Auer (1984) and Karr (1997).

Additionally all individuals participating in (any) active labour market policy measures could be counted as unemployed. The argument here is, that these individuals are looking to take up employment but were not able to do so – may be for labour market (no adequate job offers available, etc.) or individual (necessity of training, etc.) reasons.

Then all individuals could be added who are reported as ill, using information on the reason for leaving and entering registered status of unemployment. One argument to do so could be, that this is comparable to the status of employment, which is not changing due to temporarily illness.

As a last step of this second concept all individuals could be included without information on their employment status, presuming that most people are searching for employment and are willing to work – at least under good conditions –, even if they are not registered as unemployed. This also refers to the concept of hidden unemployment.

After the description of the two contrasting definitions of unemployment, the rest of this paper will demonstrate how these definitions effect results of statistical analyses. To implement these definitions, we have to first understand the structure of register data.

3. The Structure of Register Data Used in the Case Study

Any implementation of a theoretical concept into empirical analysis has to consider the structure of the data used. In this section, the structure of a merged register data set is explained by a fictitious individual employment history that could occur in real data. In particular, it illustrates how different concepts could be implemented to determine the status of unemployment and to calculate its duration.

The Integrated Employment Biographies (Integrierte Erwerbsbiographien, IEB) contain merged administrative data from Germany.¹⁰ Information is included on reported dependent employment (Source: BeH),¹¹ registered unemployment and on job search (BewA), unemployment benefits (LeH) and on participation in active labour market policies (MTH). All this information are reported as spells with start and end dates and an identifier for the individual.¹²

¹⁰ A representative 2.2 percent sample is made available for scientific use by the research data center of the FES (FDZ). For a description see Jacobebbinghaus/Seth (2007).

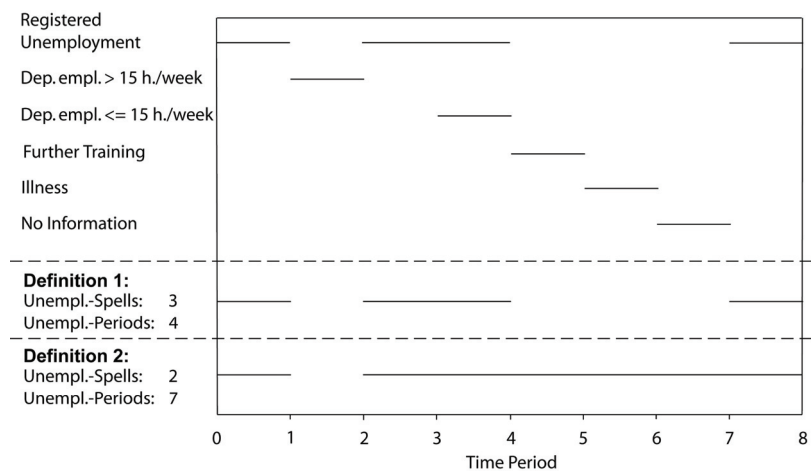
¹¹ The abbreviations on the source of the data are deducted from the pre-merged data sets: BeH: Beschäftigtenhistorik; BewA: Bewerberangebot; LeH: Leistungsempfängerhistorik; MTH: Maßnahmeteilnehmerhistorik.

¹² The merge of the several sources leads to overlapping and partially contrary information, see section 5.

Figure 1 illustrates the (partially parallel) spells of different labour market states over eight time periods. In real data these spells would probably have different lengths. Derived from the definition of registered unemployment (definition 1), the Figure shows a resulting number of three unemployment spells and a cumulated unemployment duration of four periods. The duration of the last unemployment of the individual would be reported as one period.

If we utilise the same fictitious individual employment history and implement the alternative concept (definition 2), there are three more periods counted as unemployment: a participation in an active labour market policy measure, i.e. further training, a period of illness and a period, where no information is available in the data. In this example the additional periods of unemployment fill the gap of the registered unemployment spells. The resulting number of unemployment spells therefore is reduced to two, while the cumulated duration is enlarged to seven periods. The duration of the last unemployment of the individual would now be reported as six periods.

Figure 1: Spell Structure of Data and the Counting of Unemployment:
A Fictitious Individual Employment History



4. Effects of Definitions on Statistical Results

After having described the structure of the data used for the case study, this section will test how the two definitions introduced in section 2 effect statistical results. The empirical implementation of the definition of registered unemployment as well as of the alternative definition into the Integrated Employment Biographies, which is merged from different register data, is straight forward (see also Köhler/Thomsen in this issue). But while implementing the two concepts using this data set, certain problems such as data inconsistencies

and missing interval information occurred. Data inconsistencies due to unfeasible overlapping of register information have already been extensively analysed by Jaenichen et al. (2005) and Bernhard et al. (2006). Therefore, it is highly recommend at least to check for these problems and to perform the data corrections suggested by Bernhard et al. (2006) if necessary.

In a first step the researcher has to decide which administrative information is used as the core information for an unemployment period, since BewA and LeH provide sets of unemployment information. While LeH information is related to the receipt of unemployment compensation, data from BewA informs about job seeking activities. For this reason, the source from which the unemployment information is taken has to be defined: BewA, LeH or from both (BewA+LeH). The choice can only be made according to the specific research question at hand: To get most comparable results for analysing registered unemployment only information from BewA on registered unemployment is used. For the same reason the corrections suggested by Jaenichen et al. (2005), Bernhard et al. (2006) and Waller (2008) are not implemented for this analysis. To adopt the alternative definition of unemployment, information from MTH is used additionally.

To reduce complexity, the empirical analyses is restricted to registered unemployment in 2005 based on SGB III: From the IEB a 5 percent sample of registered unemployed - as counted by the FES - was drawn at random¹³ (in the following: reference dates). Based on this sample, the duration of unemployment is computed. In a second step, the additional information is added and the duration of unemployment according to the alternative definition is calculated, based on the same sample.

Therefore, in the first step, each uninterrupted unemployment period shown by the administrative record, i.e., each uninterrupted BewA spell, is considered as unemployment. If the period immediately before this unemployment spell is an illness of a maximum length of 6 weeks, the begin day of the unemployment spell was corrected to the begin date of the illness. If the period immediately before this new begin date is another unemployment spell, the begin date is corrected again, and so on.¹⁴ Based on definition 1, the individual duration of registered unemployment then was computed as the time between the (corrected) begin date of the unemployment spell and the reference date (implementation 1).¹⁵

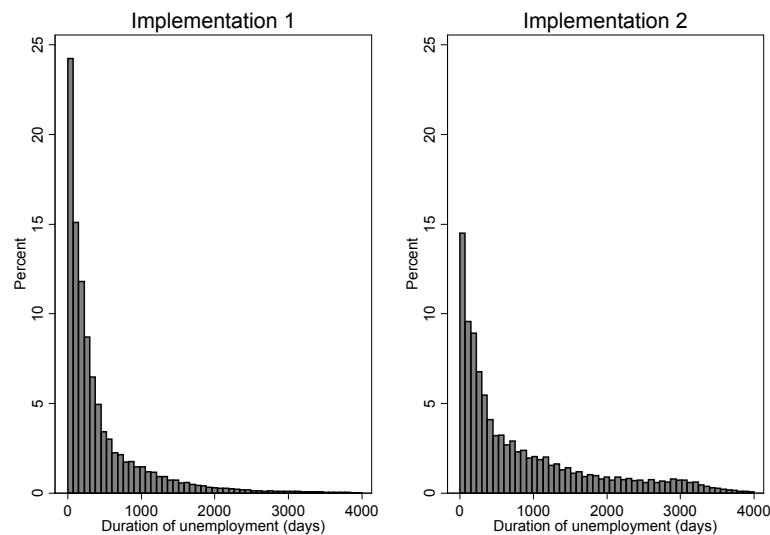
¹³ Due to the calculation of an average unemployment rate with oversample long-term unemployment (as official statistics do, see footnote 9), the drawing was split into 12 parts drawn independently at every 15th of each month 2005.

¹⁴ For comparability each interruption of up to 7 days is ignored.

¹⁵ Even if the criteria of registered unemployment are adopted most accurately, the results presented here are not comparable to unemployment durations published by the FES.

Based on definition 2, the same rules, but expanded to information from MTH about participation in active labour market measures, were applied for the calculation of the unemployment duration (implementation 2). Figure 2 shows the distribution of the duration of unemployment in days as percentage for both, implementation 1 and 2.

Figure 2: Distribution of Unemployment Duration 2005



Source: IEB, own calculations

The mean length of unemployment moves from 449 days (implementation 1) to 868 days adopting implementation 2. For interpretation the duration of unemployment can be aggregated to common groups with certain potential risks of social exclusion. Table 1 shows the shift from groups with low to those with high risk: While very short-term unemployment, short-term unemployment and the number of individuals under risk of getting long-term unemployed are lower in implementation 2, about 60 percent (implementation 1: 34 percent) of all unemployed are categorised as long-term unemployed.

Table 1: Distribution of Length of Unemployment 2005,
Using Different Implementations of Unemployment

Length of Unemployment	Implementation 1	Implementation 2
Up to 90 days	26.9 %	13.4 %
91-180 days	17.1 %	9.9 %
180-365 day	21.4 %	16.0 %
More than 1 year	34.6 %	60.7 %

Source: IEB, own calculations

5. Conclusion

Quantitative analysis - based on national register data – has to take into account, that institutional settings determine the structure of the data. This article analysed such register data derived from German Public Employment Service, comparing the duration of unemployment utilising the original underlying concept of unemployment with the implementation of an alternative definition. Starting with a legal definition (1) of registered unemployment an alternative definition (2) was discussed to exemplify the variety of theoretical driven approaches. The implementations of these approaches give an impression whether and how much definitions affect results of labour market research.

And the effect shown is strong: The mean length of unemployment moves from 449 days (implementation 1) to 868 days adopting implementation 2. This is due to a shift from groups with low to those with high risk of social exclusion: While very short-term unemployment, short-term unemployment and the number of individuals under risk of getting long-term unemployed are lower in implementation 2, about 60 percent (implementation 1: 34 percent) of all unemployed are categorised as long-term unemployed.

Definitions and implementations are depending on the theoretical framework of research undertaken. Therefore, researchers have to think about definitions and implementation very carefully in advance – before starting any quantitative analysis.

References

- Auer, Peter (1984): Maßnahmen zur Reintegration von Langzeitarbeitslosen. Discussion Paper IIM/LMP 84-20, Berlin: WZB.
- Bernhard, Sarah / Dressel, Christian / Fitzenberger, Bernd / Schnitzlein, Daniel / Stephan, Gesine (2006): Überschneidungen in der IEBS: Deskriptive Auswertung und Interpretation. FDZ Methodenreport 04/2006, Nürnberg: IAB.
- Esping-Andersen, Gøsta (1990): The Three Worlds of Welfare Capitalism, Cambridge: Polity Press.

- Jacobebbinghaus, Peter / Seth, Stefan (2007): The German Integrated Employment Biographies Sample IEBS. In: Schmollers Jahrbuch. Journal of Applied Social Science Studies, 127 (2), 335-342.
- Jaenichen, Ursula / Kruppe, Thomas / Stephan, Gesine / Ullrich, Brita / Wießner, Frank (2005): You can split it if you really want - Korrekturvorschläge für ausgewählte Inkonsistenzen in IEB und MTG. FDZ Datenreport 04/2005, Nürnberg: IAB.
- Karr, Werner (1997): Die Erfassung der Langzeitarbeitslosigkeit – Ein kaum beachtetes Meßproblem. IAB Kurzbericht (5), Nürnberg: IAB.
- Kruppe, Thomas (2003): Beschäftigungssysteme als Summe individueller Übergänge am Arbeitsmarkt. Beiträge zur Arbeitsmarkt- und Berufsforschung 271, Nürnberg: IAB.
- Kruppe, Thomas / Müller, Eva / Wichert, Laura / Wilke, Ralf A. (2008): On the definition of unemployment and its implementation in register data - the case of Germany. In: Schmollers Jahrbuch. Journal of Applied Social Science Studies, 128 (3).
- Waller, Marie (2008): On the Importance of Correcting Reported End Dates of Labor Market Programs. In: Schmollers Jahrbuch. Journal of Applied Social Science Studies, 128 (2). 213-236.
- Zimmermann, Benedicte (2006): Arbeitslosigkeit in Deutschland – Zur Entstehung einer sozialen Kategorie. Frankfurt/New York: Campus.